REVIEW
IMMUNOLOGY TODAY

HIV infection: how effective is drug combination treatment?

Zvi Grossman, Mark Feinberg, Vladimir Kuznetsov, Dimiter Dimitrov and William Paul

The rate of decline of plasma HIV RNA in patients treated with anti-retroviral drugs has been postulated to reflect the half-lives of previously HIV-infected cells. Here, Zvi Grossman and colleagues argue that the observed decline is explained by the kinetics of ongoing infection cycles. Residual cell-to-cell infection that becomes increasingly difficult to block could stabilize cellular provirus reservoirs.

Highly active anti-retroviral therapy (HAART) substantially reduces the viral load in most HIV-infected patients. It is widely believed that this results from the drugs' efficient blockade of *de novo* infection of target cells. Accordingly, most HIV that persists under HAART originates from a small number of long-lived infected cells, especially a slowly decaying pool of latently infected $CD4^+$ T cells that are converted into short-lived HIV producers when activated[1,2]. This concept of the effect of HAART is based on several observations: (1) the ability of drugs to block viral replication completely *in vitro*; (2) the rapid decline in the viral load; and (3) the apparent failure of drug-resistant mutants to occur and/or be selected in effectively treated patients.

However, the interpretation of these observations needs to be examined carefully. As we will discuss, it is likely that virus replication might continue under HAART, and that low viral loads are actively sustained through *de novo* infection events[3]. Indeed, it has been reported that unintegrated viral DNA persists in $CD4^+$ cells of treated individuals, indicating ongoing viral replication[4], as such DNA is believed to be unstable and to be found only in recently infected cells.

The extent to which virus propagation from infected to uninfected cells is blocked is of great practical significance. The answer will be important in determining whether HAART could eradicate HIV infection, and also for estimating how long treatment would need to be sustained for this to occur. If the substantial fall in plasma HIV RNA reflects only a transition to a lower but stable frequency of infection cycles, then treatment would not be expected to lead to eradication. Should persistence of HIV infection be maintained through a long-lived reservoir of latently infected $CD4^+$ T cells, elimination of this reservoir might be accomplished by use of immunological stimulation to activate these cells, thereby inducing activation-associated or virus-mediated cell death. However, for this approach to succeed, one would have to be confident that the inhibition of viral transmission is sufficiently effective that new cycles of chronic or latent infection are not induced by such activation. If they are, then externally imposed activation might enhance the level of systemic infection. Furthermore, residual virus replication and reinfection would be associated with a risk of development of drug resistance. Importantly, if low levels of ongoing replication take place during HAART, it is possible that the inherent longevity of latently infected $CD4^+$ T cells has been overestimated because the drugs used in recent studies might have failed to block regeneration of this pool completely. Therefore, treatment that blocks transmission completely might result in viral eradication more rapidly than is presently expected.

Viral load kinetics

The concept of highly efficient blockade of viral replication by effective drug treatment has derived much of its credibility from being incorporated into a simple model that successfully fits the kinetic data[1]. The main observation was a rapid fall in plasma HIV RNA of up to two orders of magnitude in the first weeks after treatment started (phase I), followed by a slower decrease in residual viral RNA (phase II). The anti-retroviral drugs used in these studies do not block virus particle production in previously infected cells; rather, they act by inhibiting *de novo* infection of target cells. It was concluded that the cells that produce most of the virus are short-lived. Assuming that the drugs blocked *de novo* infection, it was further postulated that the rate of the exponential fall in blood levels of HIV RNA during phase I essentially reflected the half-life of these cells. Similarly, the slower rate of decline in residual virus during phase II was interpreted as reflecting longer half-lives of chronically and latently infected cells[2].

Although this interpretation is consistent with the short-term kinetic data, the depiction of the death process of infected cells in terms of a 'half-life' is biologically implausible. It is equivalent to assuming that an infected cell has a fixed probability of dying at any time after being infected. More precisely, if a cell is alive at time t, the probability $p(t)$ of cell death in the interval $(t, t + dt)$, where dt is a small time increment, is given by the product kdt, where k is a constant independent of t. This is similar to the decay of radioactive atoms. However, it is more in keeping with biological processes to assume that infected lymphocytes undergo an aging-like process. That is, the death of an HIV-producing cell follows productive infection after a variable but restricted 'time delay'. If this picture of cell death represents the fate of infected cells more accurately, then the simple exponential kinetics that are seen require a different explanation.

0167-5699/98/$ – see front matter © 1998 Elsevier Science. All rights reserved.
PII: S0167-5699(98)01353-X

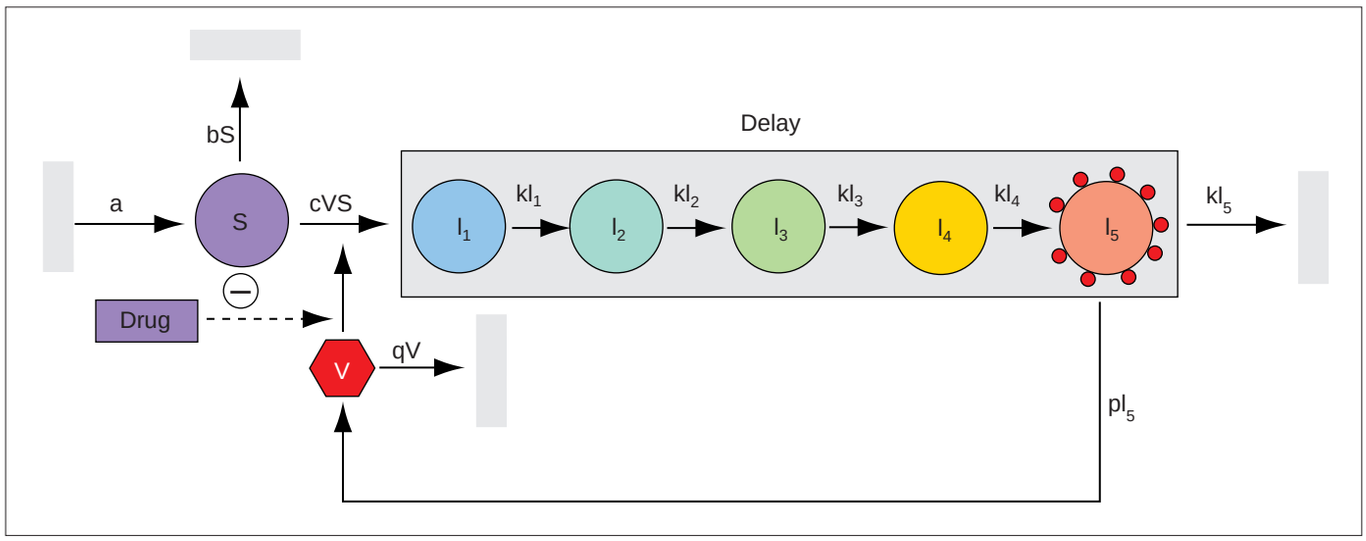

Fig. 1. *A simple, conceptual model representing cycles of infection of activated, susceptible cells (S) with delay-type kinetics of infected-cell death. Susceptible cells appear at a fixed rate, a, and are removed (or become non-susceptible) at a rate bS. They are infected by ambient virus (V) at a rate cVS, where c is an infection-rate coefficient. The delay between cell infection and death is represented as a simple multistep process (Box 1), in which the infected cell progresses, with the same transition-rate coefficient, k, from one stage to the next, with cell death and virus release occurring at the last stage. We have assumed that infected cells have a (variable) lag period and then become actively engaged in virus production and release, and in parallel become highly susceptible to cell death. The rate coefficients of virus production and decay are p and q, respectively. The drug inhibits the production of infectious virus, thereby reducing the infection-rate coefficient, c.*

Reinterpreting the exponential decline of HIV RNA

Such an explanation is provided by assuming that *de novo* infection is not blocked completely but only reduced. An exponential decline of viremia can be readily simulated. As an illustration, we used a standard model of direct HIV transmission among susceptible $CD4^+$ cells, with a delay-type kinetics of infected-cell death (Fig. 1; Box 1). A variable delay can be modeled in different ways. Here it is depicted as a multistep process in which each step is characterized by a half-life. The total delay, from infection to death, is variable but has a restricted range. If the average delay is one day, then almost all of the short-lived HIV-producing cells might die within two to three days of being infected.

The efficiency of the infection at the quasi-steady state is defined, indirectly, by the requirement that each infected cell produces during its lifetime enough virus to infect, on average, one other cell. This in turn requires that the 'basic reproduction ratio' (BRR) of the infection should be greater than 1. BRR is a measure of infectivity: the number of cells that each HIV^+ cell would infect when it is first introduced into an uninfected host. Thus, BRR is a function of the probability that a susceptible cell becomes infected, as well as the number of susceptible cells. Before treatment, BRR >1 (the actual reproduction ratio, by definition, equals 1 at steady state). If drug treatment lowers the BRR below the threshold of 1, decreasing numbers of cells are infected in each cycle so that the infection dies out. Treatment need only cause a sufficient reduction in the number of infectious virions produced per cell, or decrease the efficiency of reverse transcription in target cells, to give a BRR <1.

As a typical example, we show that the phase I kinetics observed by Perelson *et al.* in an individual treated with ritonavir (Ref. 1) can be readily fitted to such a model (Fig. 2). The new, reduced BRR is calculated by fitting the exponential portion of the experimental slope, and the delay associated with infected-cell death is estimated from the initial shoulder. Note that a reproduction ratio of 0.3 means that each 100 infected cells can still infect 30 new cells in the next round (instead of each 100 cells infecting 100 new cells at steady state) – a significant residual capacity.

By contrast, if the slope represents the death rate of infected cells[1,2], then it cannot provide information about the degree of ongoing viral transmission during anti-retroviral therapy, which was previously either neglected or considered an unspecified 'correction'. In our model, the slope is determined almost entirely by the degree to which the replication rate has been reduced by the drug. The time to death after infection has relatively little effect on the slope of HIV RNA decay; rather it accounts, at least in part, for the 'shoulder' of the kinetic curves. We emphasize that the ability to fit kinetic data using a given model proves consistency but, in general, does not validate the model. Clearly, interpretation of the data considered here must rely on plausible biological assumptions rather than on a successful curve fitting *per se*.

Testable predictions

There is a way to distinguish experimentally between the different models of cell death, *in vitro* at least. This would require efficient infection of $CD4^+$ T cells in culture and the blockade of any further infection by, for example, effective concentrations of inhibitory anti-CD4 monoclonal antibodies or anti-retroviral drugs. The kinetics of decline of viable infected cells *in vitro* would indicate whether infected-cell death is better characterized by a half-life or by a delay. Recent findings (Ref. 5; D.S. Dimitrov, unpublished) are consistent with delay-type kinetics of infected-cell death. Once cells from a T-cell line became infected and began to die, virtually all the cells died

Box 1. Model of the kinetics of plasma virus after initiation of therapy

A variable, or 'distributed', delay from productive cell infection to cell death can be modeled by dividing the process into several sequential stages, with a fixed coefficient for the (continuous) transition between them. These stages may represent real steps in the infection process or be regarded just as a mathematical trick. When the mechanism by which HIV kills cells is defined, it should permit a better description of the intermediate steps. Another way of modeling a variable delay is to introduce a distribution of delay times for example, Gaussian, around a fixed average. If the sequential stages in the state of an infected cell are denoted by I_j, where j = 1,..., n, then the equations are:

$dI_1/dt = -kI_1$
$dI_j/dt = k(I_{j-1} - I_j)$, j >1

Fitting the data points in Fig. 2 gave n = 5 and k = 2.76
With the notation defined in the legend to Fig. 1, the model's equations are:

$dS/dt = a - bS - cSV$
$dI_1/dt = cSV - kI_1$
$dI_j/dt = k(I_{j-1} - I_j)$, j = 2,..., 5
$dV/dt = pI_5 - qV$

When a steady state exists, it is given by:

$S_e = qk/pc$
$V_e = (b/c)(R_0 - 1)$, where $R_0 = apc/bqk$ is the basic reproduction ratio;
$I_{je} = (q/p)V_e$

For the simulation, we chose b = 1 days^{-1}, q = 3 days^{-1}, and different values of R_0. The exact value of q cannot be estimated from the data, but we only need to assume that the lifespan of free virus is short (<1 day). For patient 104 (Ref. 1), $V_e = 1.5 \times 10^5$. The steady-state values of infected cells, I_{je}, were normalized to 1, with no diminished generality. The post-treatment value of R, the basic reproduction ratio, and k, the time constant that defines the delay, were fitted from the data. When R_0 at the pretreatment steady state was 3, the post-treatment value was reduced to 0.29. When $R_0 = 6$, the post-treatment ratio R was estimated to be 0.28. This testifies to the robustness of our fitting procedure.

within three days. Death rates as great as 1.5 days^{-1} ($t_{1/2}$ = 0.4 days) would account for these data without invoking delays, but such rates appear to be biologically untenable because maturation of the productively infected state alone requires about one day.

Our model predicts that increasingly potent anti-retroviral therapies will yield an increasingly rapid decline of plasma viremia until a maximum is reached. Indeed, original phase I rates of decline[1], using single anti-retroviral agents, were smaller than those seen when more potent combination regimens were used[2,6]. There are now substantial additional data indicating that use of four or more anti-retroviral drugs leads to more-rapid rates of decline in plasma HIV RNA levels (Ref. 7; M.A. Polis *et al.*, unpublished). This is also consistent with studies on patients treated with ritonavir[8,9]. The slope of the exponential decline of plasma HIV RNA in these patients depended on the dose of the drug. This implies that a drug concentration that did not achieve maximum suppression of infection in a single cycle could still greatly reduce viral burden, although at a slower exponential rate. Significantly, an inverse correlation was found between the slope of decline of plasma HIV RNA and the probability of a subsequent rebound in viremia during treatment[9]. Although the conventional explanation of the rebound is the selection of drug-resistant mutants, an alternative explanation[10] is that a slower decline followed by rebound reflects a reduction in the amount of infectious virus in treated patients that is insufficient to reduce the BRR below the threshold of 1. Initially, the virus load declines rapidly as fewer virus particles are produced than before. Readjustment of the numbers of target cells to the fall in virus load may follow more slowly. Their numbers gradually increase and the virus load rises again to reach a new steady-state level that is not substantially lower than before treatment. Steeper rates of decline probably reflect a reduction of the BRR below 1, so that no rebound occurs. However, there is no need to assume that in those cases in which rebound does not occur blockade of the infection is virtually complete. The two explanations of the kinetics of HIV-rebound following ineffective treatment are not mutually exclusive. In fact, continued replication of drug-sensitive virus in patients on anti-retroviral therapy will probably predispose them to the development of drug resistance.

A direct way of testing our hypothesis would involve infecting macaques with simian immunodeficiency virus (SIV), treating them with an effective drug combination and then injecting labeled autologous target cells that had been obtained before infection, and after storage were activated and expanded. Subsequent enumeration of labeled infected cells, if any, and comparison with the

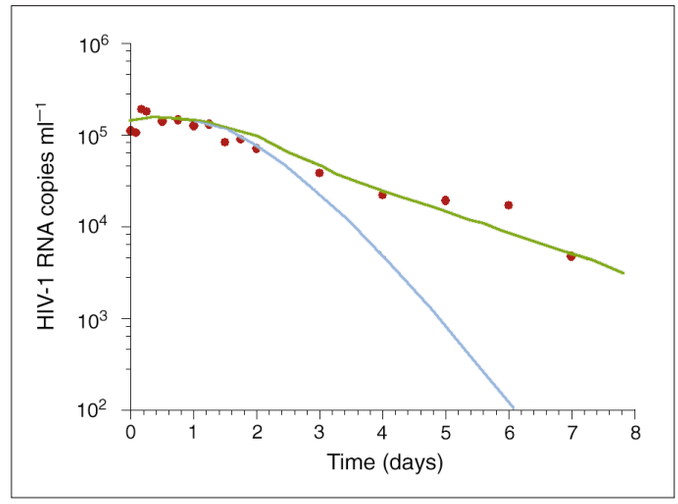

Fig. 2. *The model described in Fig. 1 is used to simulate phase I kinetics of virus concentration decline after the initiation of effective drug treatment. The data points are reproduced from Ref. 1 (patient 104). The delay is approximately one day (with a distribution generated by five sequential steps), and the fitted basic reproduction ratio of the infection after treatment is ~0.3 for various choices of other parameters that yielded equally good fits. The blue line represents the decline in virus load that would have been seen with the same delay-type kinetics of infected-cell death if* de novo *infection was blocked completely (infection rate coefficient, c = 0).*

number of labeled cells that became infected when returned to untreated SIV-infected macaques, would provide a measure of the blocking efficiency of the drugs.

We have shown that reduced efficiency of HIV infection can plausibly account for the observed kinetics. Once this is realized, the possibility that continuing replication of HIV might be sustained during HAART should be considered, taking into account that there may be more than one mode of virus transmission.

The role of chronically and latently infected cells in sustaining HIV infection

It has been argued previously[3,11] that HIV might generally not sustain itself in the body simply through continuous cycles of productive infection of $CD4^+$ T cells. In such infection cycles the transmitted virus is typically produced at a distance by T cells that have been recently infected. Activation of $CD4^+$ T cells, required for virus replication in these cells, is believed to occur in local microenvironmental domains in the form of transient bursts[12]. Thus, the continuity of productive infection cycles would depend on efficient long-range transmission of HIV from one set of domains to the other. This, in turn, would require the maintenance of sufficiently high concentrations of ambient virus across lymphoid tissues at all times. This mode, designated long-range transmission (LRT), may be the dominant mode only when viral burdens are high.

By contrast, long-lived cellular sources of HIV provide the capacity to infect newly activated cells at close range, despite the discontinuity of activation events in time and space. These sources allow the production of virus in the immediate environment in which activation occurs. Indeed, the continuity of productive infection is a systemic observation, but need not exist at the local level. Thus, virus replication might generally occur in multiple isolated episodes associated with local immune activation events. The crucial steps during such episodes are (1) virus transmission to newly activated $CD4^+$ T cells from chronically infected cells (either the antigen-presenting cells or bystanders), follicular dendritic cells (FDCs), or locally activated, latently infected cells; (2) local transmission of HIV among co-activated $CD4^+$ T cells, resulting in self-limiting amplification of virus production; and (3) infection of new cells that become chronically or latently infected, or the durable binding of HIV to new FDCs, providing opportunities for subsequent cycles of infection.

This proximal activation and transmission (PAT) mode emphasizes the lack of spatial and temporal uniformity in virus concentration throughout the tissues, and that T cells activated in the presence of a cellular source of HIV are far more likely to become infected than activated cells simply exposed to cell-free HIV in extracellular fluid[13]. This would be particularly true when viral burdens are low. An important role for PAT in sustaining the infection is not only biologically plausible[3], but also consistent with compartmentalization of HIV genotypes as revealed by *in situ* analysis[14,15]. In particular, these studies indicate that activated T cells are much more likely to be infected by cellular HIV stored locally at the site of activation than by ambient, cell-free HIV.

Thus, chronically and latently infected cells do not themselves contribute significantly to the virus load but might be instrumental in sustaining the infection. Their role is to provide the sparks that can repeatedly ignite the flames of local and transient bursts of infection. Most of the virus arising in the PAT mode is a result of amplification: once new $CD4^+$ T cells are locally activated and productively infected, they can in turn infect other activated cells in their vicinity, amplifying the initial events. These local bursts of infection do not spread systemically, unlike the acute phase of the infection, because of the early establishment of protective anergy[3,11] and, possibly, of cell-mediated immunity. Local bursts of infection restore or increase the local levels of chronic and latent infection. The relative stability of the process is probably due to the longevity of resting or activation-resistant, provirus-containing cells[16,17], and to a low frequency of new infections that result in a chronic or latent infection state.

HIV replication under HAART: the flames have been suppressed, but is the fire out?

Although the numbers of productively infected $CD4^+$ cells in patients with high viral loads are reduced dramatically by HAART, it is possible that infection continues to be actively sustained by the PAT mode. Anti-retroviral drugs might reduce the infection rate of activated $CD4^+$ lymphocytes by cell-free virus below the self-maintenance threshold, but might not reduce the efficiency of the local mode, involving relatively persistent cellular sources of the virus and direct cell-to-cell transmission, below this threshold. In particular, directed, close-range infection during cell-to-cell interaction might be less inhibitable (meaning that fewer viruses produced by a cell may be sufficient for the infection of another cell) than infection that depends on the concentration of cell-free HIV in the extracellular space. PAT is more effective because the spatial proximity to the viral source provides higher local concentrations of particles, and the time contingency of activation and virus transmission might render the target cells more susceptible and the virus more infective (cell-free HIV is believed to lose infectivity rapidly). Indeed, a quantitative temporal analysis of HIV infection in tissue cultures demonstrated that cell-to-cell transmission was much more efficient, by about 100-fold, than infection by cell-free virus[13]. In addition, anti-retroviral drugs might not gain equal access to all compartments where HIV-infected cells reside.

The large reduction of viral loads under HAART might reflect a transition from LRT to the PAT mode. Alternatively, if PAT was also the relevant mode before the initiation of treatment, the reduction might reflect a failure to amplify local cell-to-cell infection events efficiently. In both cases, a collapse to a new, low-level steady state of virus replication is expected to manifest essentially exponential kinetics at first, as in Fig. 2. A given number of virus-producing cells will cause infection of a smaller number of susceptible cells, either systemically or in several local environments, leading to the production of less and less virus. The rate of decline slows down as the new steady state is approached, since ongoing cycles of infection become increasingly difficult to block: this is because virus transmission increasingly depends on genuine PAT events and less on amplification, and because fitter viruses and more-susceptible targets, in terms of proximity or state of activation, are progressively selected.

Thus, it is possible that a low level of infection is actively sustained in some or all treated patients with undetectable plasma HIV RNA. Indeed, with the development of more-sensitive assays, the proportion of treated patients that are classified as having 'no detectable viremia' is smaller than was the case with less-sensitive assays[18]. We contend that virus production in such individuals might not be attributable solely to activation of, and virus production by, cells that had been latently infected before treatment, or to residual active infection in isolated sites that are inaccessible to the drugs. Rather, virus might also be produced by recurrent rounds of productive infection, collectively replenishing a small but stable pool of chronically and latently infected cells that can continue this cycle. Even if HIV replication is very low in treated individuals, the infrastructure of chronically and latently infected cells might still be partly or largely intact. This might explain why, after cessation of treatment, HIV usually rebounds to pretreatment levels. Indeed, sustained, low-level infection, with a reduction in systemic activation, could be associated with a risk of rebound to a level even higher than the initial baseline, at least transiently. This is because potential target cells might have 'tuned down' their activation thresholds, and thus escaped from a state of partial anergy, in which they were relatively resistant to infection[19].

By contrast, virus eradication can be approached if activation of latently infected cells does not often lead to infection of new cells, so that the frequency of infection cycles progressively diminishes. In some individuals, and with some modes of therapy, the efficiency of local activation and infection may have been reduced below the infection-maintenance threshold so that the number of infected cells might in fact be decreasing progressively.

If replication continues under effective treatment, why do resistant mutants rarely evolve? Resistance to multiple drugs requires multiple mutation steps and the mutation frequency is a function of the rate of virus replication[20,21], but the rate of replication in treated individuals is much lower than that before treatment. Under these conditions the probability of resistant mutants appearing is very low[22]. We would predict, however, that infrequent emergence of resistant mutants will occur in a population of such individuals.

Conclusion

HAART can be extremely effective in reducing HIV burden but this does not imply that it does so by actually stopping viral transmission. We argue that data showing decay of the virus in treated individuals are more consistent with variable degrees of inhibition and with infected cells dying after a time-delay rather than decaying exponentially. This interpretation allows us to estimate the residual transmission of HIV that still occurs in the presence of drugs that effectively reduce the viral load: we find such residual transmission to be far from negligible. Furthermore, we propose that HAART is more effective in blocking long-range transmission of HIV than in inhibiting proximal activation and transmission, and also that long-lived infected cells are responsible for ongoing cycles of infection in treated individuals. If this is true, then viral eradication might not be achieved by this therapeutic strategy alone. If all infected cells are to be eliminated, more effective drug therapy capable of interrupting proximal activation and transmission might be required. Alternatively, other interventions might be necessary, such as controlled activation of residual infected cells and/or the induction of effective anti-HIV immunity under the 'cover' of the reduction in viral replication induced by HAART.

We thank Russel Anderson, Tom Folks, Ashley Haase, Shmuel Livnat, Angela McLean, Steve Merrill, Robert Siliciano and Harold Varmus for their critical review of the manuscript and for their useful suggestions

Zvi Grossman, Mark Feinberg *and* ***William Paul*** *(wepaul@nih.gov) were at the NIH Office of AIDS Research, Bethesda, MD 20892, USA during the preparation of this work;* ***Vladimir Kuznetsov*** *and* ***Dimiter Dimitrov*** *are at the Laboratory of Experimental and Computational Biology, NIH, Frederick, MD 21702-1201, USA;* ***Zvi Grossman*** *is also at the Dept of Physiology and Pharmacology, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel;* ***Mark Feinberg*** *is currently at the Emory University School of Medicine, 1518 Clifton Rd, NE Atlanta, GA 30322, USA;* ***William Paul*** *is currently at the Laboratory of Immunology, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, MD 20892-1892, USA.*

References

1 Perelson, A.S., Neumann, A.U., Markowitz, M., Leonard, J.M. and Ho, D.D. (1996) *Science* 271, 1582–1586
2 Perelson, A.S., Essunger, P., Cao, Y. *et al.* (1997) *Nature* 387, 188–191
3 Grossman, Z., Feinberg, M.B. and Paul, W.E. (1998) *Proc. Natl. Acad. Sci. U. S. A.* 11, 6314–6319
4 Chun, T-W., Stuyver, L., Mizell, S.B. *et al.* (1997) *Proc. Natl. Acad. Sci. U. S. A.* 94, 13193–13197
5 Chun, T.W., Chadwick, K., Margolick, J. and Siliciano, R.F. (1997) *J. Virol.* 71, 4436–4444
6 Havlir, D.V., Eastman, S., Gamst, A. and Richman, D.D. (1996) *J. Virol.* 70, 7894–7899
7 Notermans, D.W., Goudsmit, J., Danner, S.A., de Wolf, F., Perelson, A.S. and Mittler, J. (1998) *AIDS* 12, 1483–1490
8 Danner, S.A., Carr, A., Leonard, J.M. *et al.* (1995) *New Engl. J. Med.* 333, 1528–1533
9 Mueller, B.U., Zeichner, S.L., Kuznetsov, V.A., Heath-Chiozzi, M., Pizzo, P.A. and Dimitrov, D.S. *AIDS* (in press)
10 McLean, A.R., Emery, V.C., Webster, A. and Griffiths, P.S. (1991) *AIDS* 5, 485–489
11 Grossman, Z. and Herberman, R.B. (1997) *Nat. Med.* 3, 486–490
12 Gulbranson-Judge, A. and MacLennan, I. (1996) *Eur. J. Immunol.* 26, 1830–1837
13 Dimitrov, D.S., Willey, R.L., Sato, H. *et al.* (1993) *J. Virol.* 67, 2182–2190
14 Cheynier, R., Henrichwark, S., Hadida, F. *et al.* (1994) *Cell* 78, 373–387
15 Cheynier, R., Gratton, S., Halloran, M. *et al.* (1998) *Nat. Med.* 4, 421–427
16 Wong, J.K., Hezareh, M., Gunthard, H.F. *et al.* (1997) *Science* 278, 1291–1295
17 Finzi, D., Hermankova, M., Pierson, T. *et al.* (1997) *Science* 278, 1295–1230
18 Cohen, J. (1997) *Science* 267, 1606
19 Grossman, Z. and Paul, W.E. (1992) *Proc. Natl. Acad. Sci. U. S. A.* 89, 10365–10369
20 Coffin, J.M. (1995) *Science* 267, 483–488
21 Condra, J.H., Holder, D.J., Schleif, W.A. *et al.* (1996) *J. Virol.* 70, 8270–8276
22 Richman, D.D. and Brown, A.J.L. (1997) *Nat. Med.* 3, 268–271